Using Bayesian tip-dating method to estimate divergence times and evolutionary rates

ZHANG Chi[1,2]

(1 *Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences* Beijing 100044 zhangchi@ivpp.ac.cn)
(2 *CAS Center for Excellence in Life and Paleoenvironment* Beijing 100044)

Abstract Bayesian tip dating is a recently developed method to estimate divergence times and evolutionary rates. It overcomes several drawbacks in traditional stepwise approach. However, it also requires more knowledge about statistics. This paper hierarchically explains the theory and computation in the Bayesian tip-dating approach, and divides the whole process into prior for the divergence times, prior for the evolutionary rates, model for the character changes and Markov chain Monte Carlo algorithm, which are key components in this method. The aim is to provide a general guidance for paleontologists in empirical data analyses.

Key words Bayesian tip dating, fossilized birth-death process, relaxed clock, Mk model, MrBayes

Citation Zhang C, in press. Using Bayesian tip-dating method to estimate divergence times and evolutionary rates. Vertebrata PalAsiatica. DOI: 10.19615/j.cnki.2096-9899.210516

贝叶斯支端定年法推断分异时间和演化速率

张　驰[1,2]

(1 中国科学院古脊椎动物与古人类研究所，中国科学院脊椎动物演化与人类起源重点实验室　北京 100044)
(2 中国科学院生物演化与环境卓越创新中心　北京 100044)

摘要：贝叶斯支端定年法是近些年开发的推断类群分异时间和演化速率的方法。它克服了传统分步计算的缺陷，但涉及到的统计学知识也更多。本文从贝叶斯统计计算的角度分层剖析了支端定年法的原理和计算过程，按照分异时间的先验分布、演化速率的先验分布、特征状态变化的模型和马氏链蒙特卡罗算法几个部分，叙述并讨论了定年计算中的主要模型和算法。旨在一定程度上为古生物学家分析实际数据提供参考。

关键词：贝叶斯支端定年法，石化生灭过程，宽松钟，Mk模型，MrBayes

中国科学院率先行动百人计划青年俊才(C类)和中国科学院战略性先导科技专项(B类) (编号：XDB26000000)资助。

收稿日期：2021-01-18

1 简介

推断类群的系统关系和分异时间是分支系统学分析的基础。如何合理利用化石的形态和地质年代的数据来完成此类推断一直是一个棘手的问题。传统的推断过程往往采用分步的策略。首先单独从形态数据出发，利用简约法构建类群的系统关系。这个关系只有拓扑结构而没有时间的信息，而且只代表给定时间和搜索条件下最优的结果(即最简约的树或其合意树)。然后固定这个拓扑结构，从化石年代出发，利用最小枝长法(Laurin, 2004)或等距枝长法(Brusatte et al., 2008)确定树中内部节点的分异时间。最小枝长法是把当前节点的时间往前推一个百万年作为祖先节点的时间；而等距枝长法则是把祖先节点到后代节点的时间的中点最为当前节点的时间。至于演化速率，可以通过祖先状态重建和每个树枝上特征变化的次数，结合上一步推断的时间来估计。这种分步计算的方式比较直观，因此被用于实际数据分析(Wang and Lloyd, 2016)。然而，该策略存在诸多缺陷。首先，它每一步都忽略了推断中的不确定性，包括树的拓扑结构、分异时间和祖先特征状态；其次，每一步都只利用了一部分数据信息，如建树时只用到形态特征，定年时只用到化石年代；再次，定年的方式很主观，对化石数量的增减很敏感，且不适用于现生类群；最后，整个过程缺乏一个严谨的统计学框架，无法对不同的模型假设进行检验。

近些年开发的贝叶斯支端定年法(Bayesian tip dating) (Ronquist et al., 2012; Gavryushkina et al., 2014; Zhang et al., 2016)很好地克服了上述问题。贝叶斯支端定年法把化石形态和年代数据整合在一次完整的计算过程中，能够尽可能地利用数据信息，同时考虑了树的拓扑结构、分异时间、演化速率以及化石年代的不确定性。该方法通过统计模型来描述特征的演化、类群的生灭以及化石的采样等过程，并借助相对成熟的贝叶斯统计框架和计算方法来进行参数估计和模型选择。但该方法相对来说比较复杂，需要较多的统计学知识，对古生物学家来说往往难于理解和上手，而且鲜有系统阐述支端定年法计算过程和参数意义的文献(Gavryushkina and Zhang, 2020)。本文逐层剖析支端定年法的计算过程，解释其中用到的重要模型和参数意义，旨在一定程度上为古生物学家分析实际数据提供参考。

本文首先介绍描述时间树的石化生灭过程(fossilized birth-death process)模型(Stadler, 2010), 然后介绍描述特征演化速率的宽松形态钟(relaxed clock)模型，接着介绍描述特征状态变化的Mk模型(Lewis, 2001), 再通过贝叶斯公式把上述模型联系起来，最后介绍估计参数后验分布的马氏链蒙特卡罗(Markov chain Monte Carlo, MCMC)算法。附录提供了计算中生代鸟类数据(Zhang and Wang, 2019)的MrBayes命令。

2 分异时间

时间树(timetree)代表类群的系统关系和分异时间，它的概率分布可以通过石化生灭过程(Stadler, 2010)来给出。该过程描述了从这些类群的最近共同祖先(树根)开始，分异、灭绝、采集化石和采样现生类群这一系列事件的发生，并对应于一棵完整树(图1A)。但是实际数据分析中无法推断这个完整树，只能推断和样本相关的部分，即样本树(图1B)。记每个树枝的分异速率(或叫成种率)为λ, 灭绝速率为μ, 沿每个树枝的化石采

样速率为ψ, 现生类群的采样概率(或采样比例)为ρ, 通过建立并求解一系列常微分方程，可以得到给定λ, μ, ψ, ρ时，样本时间树$T = \{\tau, \mathbf{t}\}$的概率分布，记为$P(T \mid \lambda, \mu, \psi, \rho)$, 其中$\tau$代表拓扑结构，**t**代表以百万年为单位的分异时间。

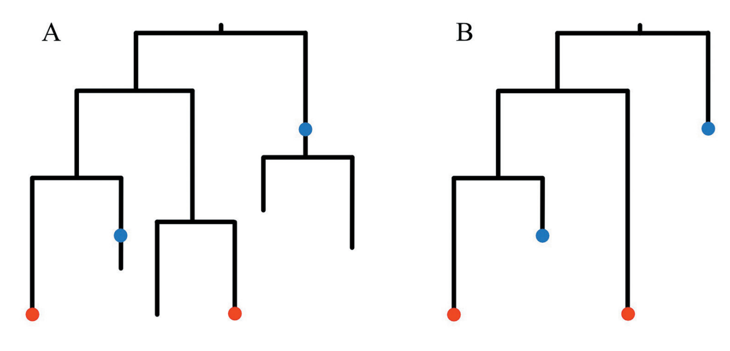


图1 石化生灭过程可能得到的时间树
A. 生灭过程对应的完整树；
B. 只保留和样本相关的分支得到的样本树，包含两个现生类群(红点)和两个化石(蓝点)
Fig. 1 Example timetree generated from the fossilized birth-death process
A. compete tree by keeping all branches; B. sampled tree by only keeping branches leading to two extant taxa (red dots) and two fossils (blue dots)

在MrBayes软件中，该生灭过程以树根时间t_1为起始条件(图2), 计算时要指定t_1的先验分布。通常这个先验比较宽泛(最大范围从0到无穷), 不过一般能从研究的类群预估一个更精确的范围，例如，其下界不会早于最古老化石的年代。化石的时间可以固定为具体的数值(百万年前), 也可以用一个均匀分布给出时间的上下界。在计算时还需要提供现生类群大致的采样比例(ρ)。现生类群可以有两种采样策略，一种是均匀随机采样(random), 另一种是多样化采样(diversity) (Zhang et al., 2016), 可以根据实际数据的情况自行选择，后者可能更符合高阶元类群的采样模式(如每个科只取一个代表的属或者每个属取一个代表物种)。对于分异、灭绝和化石采样速率，程序为了设置先验方便，重新参数化为$d = \lambda - \mu$, $v = \mu / \lambda$, $s = \psi / (\mu+\psi)$。d的默认先验为指数分布(范围从0到无穷), v和s的默认先验为均匀分布(范围从0到1, 更一般地为贝塔分布)。这样，时间树包括分异时间等参数的先验分布基本就确定了。

需特别提到的是，有些数据只包含了化石而没有现生类群。这时一般假设生灭过程在未到达现今时间点所有类群就都灭绝了。因此现

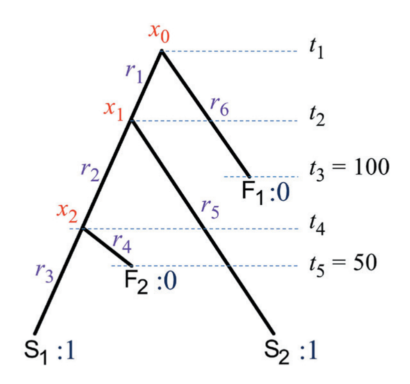


图2 用于各个概率分布公式中参数的示例树上化石F_1和F_2的特征状态为0, 现生类群S_1和S_2的特征状态为1, 内部节点的特征状态用x_0, x_1, x_2表示。
化石的年代分别为t_3 = 100 Ma, t_5 = 50 Ma, 树根的时间为t_1, 其他分异时间为t_2和t_4。记$\mathbf{t} = \{t_1, t_2, t_3, t_4, t_5\}$。各个枝上的特征演化速率为$\mathbf{r} = \{r_1, r_2, r_3, r_4, r_5\}$
Fig. 2 Example parameters and symbols used in the probability distributions
The character states for fossils F_1 and F_2 are 0, for extant taxa S_1 and S_2 are 1, for internal nodes are x_0, x_1, x_2. The ages of fossils F_1 and F_2 are t_3 = 100 Ma and t_5 = 50 Ma
The root age is t_1 and the remaining divergence times are t_2 and t_4. Denote $\mathbf{t} = \{t_1, t_2, t_3, t_4, t_5\}$. The evolutionary rates on the branches are $\mathbf{r} = \{r_1, r_2, r_3, r_4, r_5\}$

生类群不论采用何种采样策略和比例(程序默认为1.0), 都没有样本被采集到。不过由于MrBayes软件的限制，最年轻的样本总被显示为现生类群(时间为0), 因此需要把整棵树的时间轴进行相应的平移，或者说分异时间要加上年代最近的化石的年代。这只是结果显示的问题而不是程序计算的错误(后续版本将修复这个问题)。

除了石化生灭过程之外，MrBayes还为时间树提供了一个均匀分布的先验(Ronquist et al., 2012)。它没有生灭和采样这些参数，只依赖于树根时间t_1, 因此只需设置t_1的先验分布。均匀分布时常被认为是没有信息的先验，但它实际上往往带有很强的信息，在定年这个问题上，会影响对分异时间的估计。而石化生灭过程看似参数很多，但实际上其设置可以很灵活。例如，分异、灭绝和化石采样速率都可以随时间变化，在不同的时间段内各自独立(Gavryushkina et al., 2014; Zhang et al., 2016)。这可能更符合实际生物学过程，同时还能推测净成种速率和化石采样速率随时间的变化。

3 演化速率

形态特征的演化速率是指每百万年每个特征期望的变化次数。对于给定的一段时间，演化速率越快，则特征最终期望的变化次数越多。一般给每个树枝一个演化速率参数，记为**r** (图2)。时钟模型应用于形态数据时被称为形态钟模型，类比于用在分子数据时的分子钟模型。严格钟(strict clock)模型假设演化速率在各个树枝上都相同，这通常不适用于形态数据，实际分析时往往需使用宽松钟(relaxed clock)模型。宽松钟模型可以分为两类，一类为独立速率，另一类为自相关(autocorrelated)速率，区别在于**r**的概率分布P(**r**)不同。

独立速率模型假设枝上的演化速率彼此独立，它们都服从均值相同的某个概率分布。常用的概率分布包括伽马分布(Lepage et al., 2007)和对数正态分布(Drummond et al., 2006)。分布的均值也被称为基准速率(base rate), 反映平均的演化速率大小。分布的方差则反映演化速率在树枝之间变化的剧烈程度：方差较小时各个速率相差不大，这意味着演化速率在整棵树上没有明显的差异；而方差越大，不同树枝上速率的差异越明显。

自相关速率模型假设后代树枝上的演化速率依赖于临近祖先那枝上的速率(例如r_2和r_5都依赖r_1, r_3和r_4都依赖r_2)。当前树枝上的速率一般假设服从对数正态分布(Kishino et al., 2001; Thorne and Kishino, 2002), 其均值为临近祖先节点的速率。同理，分布的方差也反映演化速率在树枝之间变化的剧烈程度。

这两类速率模型往往对分异时间的估计也有影响，这主要是因为自相关速率模型会倾向于速率的变化是渐进的，而独立速率模型没有这种限制，会更适应临近树枝间速率变化比较剧烈的情况。对化石形态数据来说，独立速率模型可能更适用。

默认的情况下，形态数据矩阵中所有特征都共享每个枝上的演化速率，因此，这个速率代表的是所有特征的平均情况。如果需要考虑不同特征演化速率的异质性，就需要对特征进行分区。一般可以按照不同特征类型或不同身体部位或功能来分，每个分区内的特征共享一组演化速率，而分区之间特征演化速率的模式是独立的，这样就可以推断不同部位或功能相关特征随时间会发生怎样的变化(Lee, 2016; Zhang and Wang,

2019)。需注意的是，分区越多，每个分区内的特征数量就越少，因此能够估计演化速率参数的信息就越少，会造成方差很大甚至参数个数超过特征数量导致无法进行参数估计。因此，要在考虑演化速率异质性和分区数量之间做一个权衡。

4 特征状态变化

有了分异时间和演化速率，就可以计算在给定时间段t和演化速率r的情况下，形态特征从一个状态变为另一个状态的概率(称为转移概率)。这个概率由Mk模型(Lewis, 2001)给出。Mk模型是描述特征状态变化最简单的模型，它假设状态之间转换的速率是相等的。这里只以两个状态的特征为例，用$P_{00}(r, t)$表示状态0保持不变的概率，$P_{01}(r, t)$表示从0变到1的概率，$P_{10}(r, t)$表示从1变到0的概率，$P_{11}(r, t)$表示状态1保持不变的概率，则$P_{00}(r, t) = P_{11}(r, t) = 1/2 + 1/2 \times \exp(-2rt)$, $P_{01}(r, t) = P_{10}(r, t) = 1/2 - 1/2 \times \exp(-2rt)$。从公式中可以发现，时间$t$和速率$r$总是以乘积的形式出现，因此，在没有化石年代信息时，两者是不可识别的。换句话说，单单依靠形态数据来建树，树的枝长为时间和速率的乘积，即距离，以每个特征期望的变化次数为单位。只有同时利用化石形态和年代数据才能将分异时间和演化速率单独估计出来。

以图2为例，化石F_1和F_2的特征状态为0, 现生类群S_1和S_2的特征状态为1, 内部节点的特征状态未知用x_0, x_1, x_2表示。F_1和F_2的时间分别为100 Ma和50 Ma。那么根据Mk模型，给定时间树$T = \{\tau, \mathbf{t}\}$和速率$\mathbf{r}$时，特征状态列0011的概率为

$$P(0011|T, \mathbf{r}) = \sum_{x_0}\sum_{x_1}\sum_{x_2} P_{x_0x_1}(t_1-t_2, r_1)P_{x_1x_2}(t_2-t_4, r_2)P_{x_21}(t_4, r_3)\, P_{x_20}(t_4-t_5, r_4)P_{x_11}(t_2, r_5)P_{x_00}(t_1-t_3, r_6)$$

其中西格玛符号代表对特征在内部节点所有可能状态的求和。由于形态特征矩阵往往只包含可变的特征，因此这个概率还要除以所有可变状态的概率，即$P(0011 \mid T, \mathbf{r})/[1 - P(0000 \mid T, \mathbf{r}) - P(1111 \mid T, \mathbf{r})]$。带有这一校正的Mk模型称为Mkv模型(Lewis, 2001)。

假设形态矩阵中的特征都彼此独立，那么就可以计算每一列特征在树上的概率，再把这些概率乘起来。这个概率被称为似然函数，表示为$P(D \mid T, \mathbf{r})$, 其中D代表形态特征矩阵数据。

5 贝叶斯公式

在统计推断时，参数都是未知的随机变量，需要根据数据来估计它们的分布，即计算$P(T, \mathbf{r}, \theta \mid D)$, 该分布称为后验分布，其中$T = \{\tau, \mathbf{t}\}$为时间树，$\mathbf{r}$为演化速率，$\theta$代表其它参数(包括$\lambda$, μ, ψ等)。根据分层贝叶斯公式，可得$P(T, \mathbf{r}, \theta \mid D) = P(D \mid T, \mathbf{r})\, P(\mathbf{r})\, P(T \mid \theta)\, P(\theta)/P(D)$。等号右侧分子中，第一项似然函数在第四节给出，第二项演化速率的先验分布在第三节给出，第三项和第四项为时间树及其参数的先验分布在第二节给出。这样公式分子中各项都可以计算了。分母$P(D)$是特征数据的概率，这需要计算对所有参数的多重积分，实际上基本无法给出解析表达式，只能通过数值算法进行近似。所以贝叶斯计算在绝大多数情况下会使用马氏链蒙特卡罗算法。

6 马氏链蒙特卡罗算法

马氏链蒙特卡罗(MCMC)算法通过构造马尔科夫链，使其平稳分布为要估计的后验分布。这里为了简化，只以一维参数的情形为示例(图3)。实际分析中，参数一般是多维的(如τ, **t**, **r**, θ), 不过算法的思想是类似的。

MCMC采用的Metropolis-Hastings算法(Metropolis et al., 1953; Hastings, 1970)可分为如下几步：

a. 为θ设定任意初始值；

b. 基于θ当前的值，建议一个新的值θ', 例如$\theta' \sim \text{uniform}(\theta - w/2, \theta + w/2)$;

c. 如果$\pi(\theta') > \pi(\theta)$, 就接受$\theta'$; 否则，以概率$\alpha = \pi(\theta') / \pi(\theta)$ 接受θ';

d. 如果θ'被接受，就更新$\theta = \theta'$;否则，保持θ不变；

e. 记录θ的值，回到步骤b。

注意到，在计算概率α的时候，会计算后验分布的比，这样后验分布分母的部分就约掉了，只剩分子部分的比。也就是说，只要能够把分子部分写出解析表达式，MCMC算法就可以使用来估计参数的后验分布了。

计算结束后，就收集到一些θ的样本。由于参数的初始状态往往比较差，MCMC需要经过很多代才收敛到后验概率密度比较高的地方，因此在估计后验分布的时候会舍弃初始的一些样本(burn-in), 只用MCMC链收敛后记录的样本来估计后验分布。MrBayes默认舍弃前25%的样本。同时，MCMC链还要迭代足够多次，以保证有足够多的有效样本来估计参数的后验分布。一般需要有效样本大小(ESS)大于100。

实际运算中，最好独立运行至少两次MCMC, 以确保两次的结果是一致的。有时链长不够，或者不同的运算卡在不同的后验分布区域，都会导致估计的结果不一致。这时调整MCMC的设置或者改善模型都可能帮助MCMC算法发挥更好的效能。使用Metropolis-coupled MCMC也是有效跨越多峰分布的手段(Lakner et al., 2008)。该算法同时运行多条MCMC链，一条为冷链(cold chain), 其余为热链(hot chains), 热链和冷链之间可以相互交换。当然MCMC样本只从冷链中采集，热链只是用来帮助跨越多峰的。MrBayes默认同时独立运行两次，每次运算使用四条链，其中一条为冷链，其余三条为热链。

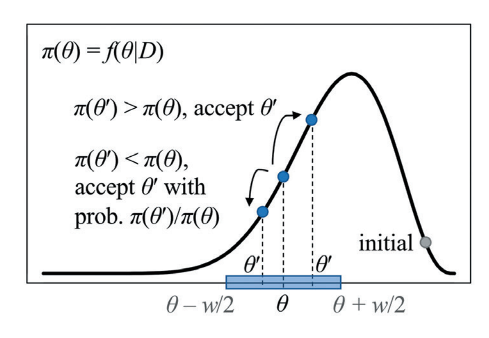


图3 马氏链蒙特卡罗算法在一维情形的示例
参数θ的后验分布$\pi(\theta)$为估计的目标(曲线)

Fig. 3 Illustration of the MCMC algorithm for one dimensional parameter
The posterior distribution $\pi(\theta)$ of parameter θ is the target distribution to be estimated (curve)

7 讨论

本文从贝叶斯统计计算的角度分层剖析了支端定年法的原理和计算过程。贝叶斯后验分布包含先验分布和似然函数，其中先验分布的两个重要组成部分：分异时间和演化速率模型，在定年分析中尤为关键，是影响定年准确性的主要因素。

石化生灭过程作为描述类群分异、灭绝和采样的随机过程，具有较大的灵活性。不过该模型还有待完善之处。现生类群的采样方式可以是随机的或多样化的，这两种情形可能都是极端，真实的采样模式可能介于两者之间，或者有的支系是随机采样，有的支系是多样化采样，但目前还没有模型能够支持这种情况。在更好的模型被开发出来之前，可能只能调整数据来尽量符合其中一种采样策略。这种处理方式在现生类群比较少或根本没有时一般问题不大。另外，分异速率、灭绝速率和化石采样速率可以按分段的方式随时间变化，不过在同一时间段内，所有树枝都共享同一速率。对于不同支系明显受到的选择压力不同或化石保存的完整性明显不同等情况，按支系分段而非时间分段的模型(Barido-Sottani et al., 2020)可能更合适，虽然这一类模型本身也有其他限制。不论怎样，石化生灭过程只是作为时间的先验分布，当数据量比较大时(包括化石在树上的分布程度和数量以及形态特征的数量和完整度), 数据在推断中会起主导作用而先验的影响减少。但是实际情况往往比较复杂，数据量也不尽如人意，这时考察不同先验的影响就尤为重要。

演化速率的先验，即形态钟模型，也会和时间相互作用，从而影响对最终分异时间的估计。这种影响在化石较少或化石在树上分布很不均时尤为明显。这主要是因为化石形态数据只提供了距离的信息(每个特征期望的变化次数), 其为时间和速率的乘积(见第四节)。当缺少化石时，就没办法准确提供时间的信息，那么对于同样的距离，可以是很长时间速率很慢，也可以是很短时间但速率很快，具体是怎样就只能取决于时间和演化速率的先验了。对于某些支系时间估计得明显偏大或偏小，但又没有化石来校正的情况，可以通过添加内部节点的校正分布来得到更合理的时间估计(O’Reilly and Donoghue, 2016)。在完全没有化石只有现生类群时，节点定年法(另一类型定年方法)(Yang and Rannala, 2006)就是通过使用内部节点的校正分布来完成的。

描述形态特征状态变化的模型也有很大的改进空间，其中涉及更多的建模和随机模拟等工作，不是本研究的重点，这里只简单讨论一下Mk模型对定年可能的影响。Mk模型假设特征各个状态间转变的速率都是相等的。这种转变有无序和有序之分(只对三个及以上状态的特征)。无序是指特征可以直接从一个状态变为任意其他状态(如从0直接变为3), 而有序是指特征只能在临近状态间直接变化(如从0到1, 从1到2, 再从2到3)。显然，有序需要更多次变化(也就是更长的距离)才能从当前状态变为不相邻的状态，因此对有序的特征假设无序的变化会低估距离。更复杂的情况是，各个状态间转变的速率未必相等甚至相差很多，极端情况像Dollo特征甚至是不可逆的。这时使用Mk模型也会造成距离估计的偏差。在计算时还假设不同特征之间都是独立的，如果有些特征有较强的相关性，则会导致演化距离的高估。前面提到，距离是时间和速率的乘积，因此在化石(时间信息)很丰富的理想情况下，距离的偏差会主要体现到演化速率上而对分异时间影响较小。但是分析实际数据时会更复杂一些，要具体问题具体分析。相关的工作还较少(Klopfstein et al., 2019), 需要更多后续研究来更详细地考察。

最后提到贝叶斯计算使用的MCMC算法。该算法的策略与简约法和似然法有明显不同。简约法寻找的是简约树长最小的树，似然法寻求的是似然值最大时参数的估

计(即最大似然树)。因此在设计树的搜索方法时，只要尽可能快速地找到最优的树就可以了。MCMC算法是为了估计参数的后验分布，这不仅仅是一个点，而是参数的空间。因此MCMC算法的效率涉及到收敛(convergence)和混合(mixing)两部分。收敛是指MCMC达到分布概率密度高的区域，混合是指MCMC能够按概率分布进行取样。提高收敛速度相对容易，可以通过如简约树长向导的方式来快速找到概率大的树(Zhang et al., 2020)。提高混合则更困难，需要设计更好的建议(proposal)方法，这是贝叶斯计算的重点也是难点。

总之，贝叶斯支端定年法作为整合的分析方法，能够结合化石形态和年代数据以及现生类群的形态和分子数据来推断类群的系统关系，分异时间和演化速率，同时考虑了树的拓扑结构、分异时间、演化速率以及化石年代的不确定性。但该方法仍处于发展初期，模型和算法的诸多方面还亟待完善，因此还有很多工作需要做。

References

Barido-Sottani J, Vaughan T G, Stadler T, 2020. A Multi-Type Birth-Death model for Bayesian inference of lineage-specific birth and death rates. Syst Biol, 69: 973–986

Brusatte S L, Benton M J, Ruta M et al., 2008. Superiority, competition, and opportunism in the evolutionary radiation of dinosaurs. Science, 321: 1485–1488

Drummond A, Ho S, Phillips M et al., 2006. Relaxed phylogenetics and dating with confidence. Plos Biol, 4: e88

Gavryushkina A, Welch D, Stadler T et al., 2014. Bayesian inference of sampled ancestor trees for epidemiology and fossil calibration. Plos Comput Biol, 10: e1003919

Gavryushkina A, Zhang C, 2020. Total-Evidence Dating and the Fossilized Birth-Death Model. In: Ho S Y W ed. The Molecular Evolutionary Clock. Switzerland: Springer Nature. 175–193

Hastings W K, 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika, 57: 97–109

Kishino H, Thorne J L, Bruno W J, 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol Biol Evol, 18: 352–361

Klopfstein S, Ryer R, Coiro M et al., 2019. Mismatch of the morphology model is mostly unproblematic in total-evidence dating: insights from an extensive simulation study. bioRxiv, 679084

Lakner C, van der Mark P, Huelsenbeck et al., 2008. Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics. Syst Biol, 57: 86–103

Laurin M, 2004. The evolution of body size, Cope's rule and the origin of amniotes. Syst Biol, 53: 594–622

Lee M S Y, 2016. Multiple morphological clocks and total-evidence tip-dating in mammals. Biol Lett, 12: 20160033

Lepage T, Bryant D, Philippe H et al., 2007. A general comparison of relaxed molecular clock models. Mol Biol Evol, 24: 2669–2680

Lewis P O, 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst Biol, 50: 913–925

Metropolis N, Rosenbluth A W, Rosenbluth M N et al., 1953. Equation of state calculations by fast computing machines. J Chem Phys, 21: 1087–1092

O'Reilly J E, Donoghue P C J, 2016. Tips and nodes are complementary not competing approaches to the calibration of molecular clocks. Biol Lett, 12: 20150975

Ronquist F, Klopfstein S, Vilhelmsen L et al., 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. Syst Biol, 61: 973–999

Stadler T, 2010. Sampling-through-time in birth-death trees. J Theor Biol, 267: 396–404

Thorne J L, Kishino H, 2002. Divergence time and evolutionary rate estimation with multilocus data. Syst Biol, 51: 689–702

Wang M, Lloyd G T, 2016. Rates of morphological evolution are heterogeneous in Early Cretaceous birds. Proc Royal Soc B Biol Sci, 283: 20160214

Yang Z, Rannala B, 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol, 23: 212–226

Zhang C, Huelsenbeck J P, Ronquist F, 2020. Using parsimony-guided tree proposals to accelerate convergence in Bayesian phylogenetic inference. Syst Biol, 69: 1016–1032

Zhang C, Stadler T, Klopfstein S et al., 2016. Total-evidence dating under the fossilized birth-death process. Syst Biol, 65: 228–249

Zhang C, Wang M, 2019. Bayesian tip dating reveals heterogeneous morphological clocks in Mesozoic birds. Roy Soc Open Sci, 6: 182062